(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
4 January 2007 (04.01.2007)

PCT

(10) International Publication Number
WO 2007/002204 A2

(51) International Patent Classification:
G01N 33/50 (2006.01)

(21) International Application Number:
PCT/US2006/024157

(22) International Filing Date: 20 June 2006 (20.06.2006)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/692,816 21 June 2005 (21.06.2005) US

(71) Applicant *(for all designated States except US)*: **THE TRUSTEES OF COLUMBIA UNIVERSITY IN THE CITY OF NEW YORK** [US/US]; West 116th Street And Broadway, New York, NY 10027 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **JU, Jingyue** [US/US]; 267 Marietta Street, Englewood Cliffs, NJ 07632 (US). **WU, Jian** [CN/US]; 362 Riverside Drive #5A9, New York, NY 10025 (US). **KIM, Dae, H.** [US/US]; 100 Morningside Drive #5h, New York, NY 10027 (US).

(74) Agent: **WHITE, John, P.**; COOPER & DUNHAM LLP, 1185 Avenue Of The Americas, New York, NY 10036 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LV, LY, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, LV, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 2007/002204 A2

(54) Title: PYROSEQUENCING METHODS AND RELATED COMPOSTIONS

(57) Abstract: This invention provides methods for pyrosequencing and compositions comprising 3'-O- modified deoxynucleoside triphosphates .

Applicants: Jingyue Ju
Serial No.: 10/521,206
Filed: November 9, 2006
Exhibit 6

PYROSEQUENCING METHODS AND RELATED COMPOSITIONS

This application claims benefit of U.S. Provisional Application No. 60/692,816, filed June 21, 2005, the contents of which are hereby incorporated by reference. The invention disclosed herein was made with government support under a grant from the Center for Excellence in Genomic Science Grant No. P50 HG002806. Accordingly, the U.S. Government has certain rights in this invention.

Throughout this application, various publications are referenced in parentheses by number. Full citations for these references may be found at the end of the specification immediately preceding the claims. The disclosures of these publications in their entireties are hereby incorporated by reference into this application to more fully describe the state of the art to which this invention pertains.

Background of the Invention

Pyrosequencing is a method based on the detection of the pyrophosphate group that is generated when a nucleotide is incorporated in a DNA polymerase reaction [1]. Each of the four deoxynucleotides (dNTPs) is added sequentially to the DNA template to be sequenced with a cocktail of enzymes and substrates in addition to the usual polymerase reaction components. If the added nucleotide is complementary with the first available base on the template, the nucleotide will be incorporated and a pyrophosphate will be released. The released pyrophosphate is converted to ATP by

sulfurylase, and this ATP is the substrate for a luciferase, e.g. firefly luciferase, which reaction produces visible light. If the added nucleotide is not incorporated, no light will be produced and the nucleotide will simply be degraded by the enzyme apyrase. This pyrosequencing technique, schematized in Fig. 1, has been applied to single nucleotide polymorphism (SNP) detection and other applications [2].

There are, however, inherent difficulties in the traditional pyrosequencing method for determining the number of incorporated nucleotides in homopolymeric regions (e.g. a string of several T's in a row) of the template. Moreover, dATP greatly interferes with the luciferase detection system, which is deficient in the detection of dATP.

Summary of the Invention

This invention provides a method for determining the nucleotide sequence of a single-stranded DNA comprising performing the following steps for each nucleic acid residue of the DNA whose identity is to be determined:

(a) contacting the DNA under DNA polymerization-permitting conditions with (i) a 3'-O-blocked dNTP selected from the group consisting of 3'-O-blocked dATP, 3'-O-blocked dCTP, 3'-O-blocked dGTP, and 3'-O-blocked dTTP, and (ii) 9°N DNA polymerase (exo-) A485l/Y409V or another DNA polymerase able to incorporate 3'-O-blocked dNTPs;

(b) (i) determining whether pyrophosphate is generated as a result of step (a), whereby (1) pyrophosphate generation indicates that polymerization has occurred and the identity of the nucleic acid residue in the DNA is that which is complementary to the 3'-O-blocked dNTP used in part (i) of step (a), and (2) the absence of pyrophosphate generation indicates that the identity of such nucleic acid residue is not that which is complementary to such 3'-O-blocked dNTP, and (ii) if pyrophosphate is not generated, repeating step (a) once, twice or three times as necessary, wherein in each repetition a 3'-O-blocked dNTP is used which is different from any 3'-O-blocked dNTP already used, and determining, after each repetition of step (a), whether pyrophosphate is generated, such generation indicating that polymerization has occurred and the identity of the nucleic acid residue in the DNA is that which is complementary to the 3'-O-blocked dNTP used in part (i) of the repeated step (a); and

(c) removing from the 3′-O-blocked dNTP polymerized in step (a) or (b), whichever is applicable, the moiety blocking the 3′-O atom of the dNTP, with the proviso that such removing step is optional in the event that there remains no further nucleic acid residue of the DNA whose identity is to be determined.

Brief Description of the Figures

Figure 1: Schematic of pyrosequencing in solution. Reactants not shown are APS, which with PPi is converted to ATP and SO_4^{2-} by ATP-sulfurylase. In addition luciferase acts on ATP, luciferin and O_2 to give AMP, PPi, oxyluciferin, CO_2 and light, and apyrase converts ATP and dNTP to AMP, dNMP and 2Pi.

Figure 2: 3'-O-Allyl-dNTP (A, C, G, T), instead of dNTP, is used in the single base extension on a solid surface. Four Allyl-dNTPs are added iteratively. Once the complementary base is incorporated, the pyrophosphate that is produced from the reaction can be detected by its reaction with the light-generating luciferase system or a receptor-indicator (R-I) coordination compound via the release of the fluorescent indicator molecule. Then the extended primers can be deallylated, washed and reused in the next round. The use of an allyl-group solves inherent problems of traditional pyrosequencing.

Figures 3A-3C: Schematic representation and step-by-step MALDI-TOF MS results for the deallylation of an allyl-modified oligonucleotide (SEQ ID NO:1) and the use of the deallylated DNA product as a primer in a polymerase extension reaction. (A) Peak at m/z 5871 corresponding to the HPLC-purified 3'-allyloxy 19-mer oligonucleotide. (B) Peak at m/z 5831 corresponding to the above DNA product without the allyl group, obtained after

30 secs of incubation with the Na_2PdCl_4 catalyst and the TPPTS ($P(PhSO_3Na)_3$) ligand at 70 °C. (C) Peak at m/z 6535 corresponding to the extension of the deallylated DNA product by Biotin-ddGTP using Thermo Sequenase DNA Polymerase.

Figure 4: Synthesis of a 3'-O-allyl-modified oligonucleotide.

Figures 5A & 5B: MALDI-TOF MS spectra showing the incorporation of 3'-O-allyl-modified dTTP into a growing DNA strand by 9°N Polymerase (exo-) A485L/Y409V; (A) unextended primer at m/z 5526; (B) primer extended with 3'-O-allyl-dTTP at m/z 5869.

Figures 6A & 6B: (A) Receptor (R): Zn^{2+}-dipicolylamine (Zn2+DPA); (B) Indicator (I): fluorescent molecule (coumarin-derived indicator). See [5].

Figures 7A & 7B: When R is titrated into I, the fluorescence intensity of I will decrease: (A) Zn^{2+}DPA is titrated into (10μM) I; (B) PPi is added to R-I coordination compound.

Figure 8: Ronaghi's real-time pyrosequencing.

Figure 9: Improved real-time pyrosequencing method.

Figure 10: Structures of four reversibly-blocked nucleotides.

Figure 11: Mass spectrometry traces showing incorporation of four different reversibly-blocked allyl-dNTPs into a growing DNA strand in the solution phase.

Figure 12: Polymerase extension reaction with 3'-O-allyl-dGTP-allyl-biodipy-FL-510 as a reversible terminator of SEQ ID NO:2.

Figure 13: Experimental results of pyrosequencing a DNA template (SEQ ID NO:3) in solution with allyl-dGTP and comparison with 'regular' unblocked nucleotides. The results indicate that allyl-dGTP is a good terminator in solution phase, and the incorporation Signal can be easily detected.

Figure 14: An experimental scheme of a method employing allyl-dGTP for pyrosequencing with attachment of the primer (SEQ ID NO:4 and SEQ ID NO:5) to a solid surface/bead using an NHS ester.

Figure 15: Comparison of pyrosequencing using 'regular' dNTPs and pyrosequencing using reversibly-blocked dNTPS (SEQ ID NO:3).

Figure 16: Pyrosequencing data using reversible terminators on sepharose bead immobilized looped primer-DNA (SEQ ID NO:3).

Figure 17: Light production by luciferase in the presence of dATP and in the presence of allyl-

dATP, demonstrating that allyl-dATP is not a luciferase substrate.

Figure 18: Technique of immobilizing double-stranded DNA (SEQ ID NO:6 (top strand) and SEQ ID NO:7 (lower strand)) to a derivatized bead and pyrosequencing using "normal" nucleotides.

Figure 19: Pyrosequencing on sepharose bead-immobilized DNA (SEQ ID NO:6 (top strand) and SEQ ID NO:7 (lower strand)) using Allyl-dNTPs.

Detailed Description of the Invention

Definitions

As used herein, and unless stated otherwise, each of the following terms shall have the definition set forth below.

PPi - pyrophosphate
dNTP - deoxynucleoside 5'-triphosphate - also known as a deoxynucleotide
APS - adenosine 5'-phosphosulfate
ATP - adenosine 5'-triphosphate
dATP - deoxyadenosine 5'-triphosphate
THF - tetrahydrofuran
TEAB - tetraethylammonium bromide
TPPTS - tri sodium salt of tri (m-sulfophenyl)-phosphine

"Nucleic acid" shall mean any nucleic acid molecule, including, without limitation, DNA, RNA and hybrids thereof. The nucleic acid bases that form nucleic acid molecules can be the bases A, C, G, T and U, as well as derivatives thereof. Derivatives of these bases are well known in the art, and are exemplified in PCR Systems, Reagents and Consumables (Perkin Elmer Catalogue 1996-1997, Roche Molecular Systems, Inc., Branchburg, New Jersey, USA).

Embodiments of the Invention

Disclosed here is a method that solves the problems of homopolymeric regions and dATP interference by using 3'-O-allyl-nucleotides as reversible terminators in pyrosequencing using either a new PPi detection system

(Chemosensing Ensemble), or the traditional luciferase detection technique.

Specifically, this invention provides a method for determining the nucleotide sequence of a single-stranded DNA comprising performing the following steps for each nucleic acid residue of the DNA whose identity is to be determined:

(a) contacting the DNA under DNA polymerization-permitting conditions with (i) a 3'-O-blocked dNTP selected from the group consisting of 3'-O-blocked dATP, 3'-O-blocked dCTP, 3'-O-blocked dGTP, and 3'-O-blocked dTTP, and (ii) 9°N DNA polymerase (exo-) A485l/Y409V or other DNA polymerase;

(b) (i) determining whether pyrophosphate is generated as a result of step (a), whereby (1) pyrophosphate generation indicates that polymerization has occurred and the identity of the nucleic acid residue in the DNA is that which is complementary to the 3'-O-blocked dNTP used in part (i) of step (a), and (2) the absence of pyrophosphate generation indicates that the identity of such nucleic acid residue is not that which is complementary to such 3'-O-blocked dNTP, and (ii) if pyrophosphate is not generated, repeating step (a) once, twice or three times as necessary (i.e. until pyrophosphate is generated), wherein in each repetition a 3'-O-blocked dNTP is used which is different from any 3'-O-blocked dNTP already used, and determining, after each repetition of step (a), whether pyrophosphate is generated, such generation indicating that polymerization has occurred and the identity of the nucleic acid residue in the DNA is

that which is complementary to the 3′-O-blocked dNTP used in part (i) of the repeated step (a); and

(c) removing from the 3′-O-blocked dNTP polymerized in step (a) or (b), whichever is applicable, the moiety blocking the 3′-O atom of the dNTP, with the proviso that such removing step is optional in the event that there remains no further nucleic acid residue of the DNA whose identity is to be determined.

The identity of a nucleic acid residue in the DNA being sequenced is that which is complementary to the 3′-O-blocked dNTP incorporated, i.e. such identity is determined by the well-established complementary base-pairing rules. For example, if a 3′-O-blocked dATP is incorporated, then the corresponding nucleic acid residue in the DNA being sequenced is a thymine. If a 3′-O-blocked dGTP is incorporated, then the corresponding nucleic acid residue in the DNA being sequenced is a cytosine, and so forth with the understanding that adenine and thymine are complements of each other, and guanine and cytosine are complements of each other. In addition, uridine is a complement of adenine.

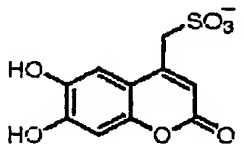
A 3′-O-blocked deoxynucleotide is a deoxynucleotide having attached to the 3′ oxygen of its sugar component a chemical group, for example an allyl group, that precludes further polymerization from the 3′ oxygen until that blocking group is removed.

This invention further provides the instant method, wherein determining whether pyrophosphate generated in step (b)(i) is performed by detecting light generated by a luciferase-based reaction. In one embodiment, the luciferase is

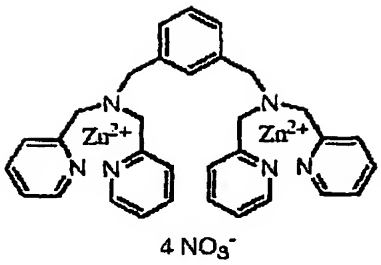
firefly luciferase. In another embodiment, the luciferase-based reaction comprises contacting the pyrophosphate with a sulfurylase under conditions permitting the generation of ATP from the pyrophosphate, and contacting the ATP so generated with a luciferase under conditions permitting the generation of light by the luciferase in the presence of ATP. A luciferase-based reaction includes, for example, the reaction of luciferin and ATP in the presence of luciferase and O_2, whereby oxyluciferin, AMP, PPi, CO_2, and light are produced. The light produced can be measured by any standard photometry technique including, but not limited to, photomultiplier, video, CCD, CCCD, and the naked eye.

In a preferred embodiment, the moiety blocking the 3'-O atom of the dNTP is an allyl moiety and the single-stranded DNA is immobilized to a solid substrate.

In another embodiment, determining whether pyrophosphate is generated in step (b)(i) is performed by detecting dissociation of a coumarin-derived indicator from a complex between the indicator and a bis-Zn^{2+}-dipicolylamine coordination compound, wherein the coumarin-derived indicator has the following structure:



and the bis-Zn^{2+}-dipicolylamine coordination compound, when in association with the coumarin-derived indicator, has the following structure:



In the preferred embodiment, the moiety blocking the 3'-O atom of the dNTP is an allyl moiety and the single-stranded DNA is immobilized to a solid substrate. In another embodiment the moiety blocking the 3'-O atom of the dNTP is an allyl moiety or a methoxymethyl moiety. Preferably, the moiety is an allyl moiety.

In the preferred embodiment of the instant method, the DNA is immobilized on a solid substrate. In different embodiments, the DNA is bound to the solid substrate via an azido linkage, an alkynyl linkage, a 1,3-dipolar cycloaddition linkage, or biotin-streptavidin interaction. The solid substrate can be, for example, in the form of a chip, a bead, a well, a capillary tube, or a slide. Also, for example, the solid substrate can be gold, quartz, silica, or plastic. In one embodiment of this invention, the solid substrate is porous.

Single-stranded DNA can be immobilized on a solid surface, for example a glass surface, by a 1,3-dipolar cycloaddition reaction in the presence of a Cu(I) catalyst. The DNA is labeled with an azido group at the 5' end, while the glass surface is modified by an alkynyl group. After the 1,3-dipolar cycloaddition between the azido and the alkynyl group in the presence of a Cu(I) catalyst at room temperature, the DNA is covalently attached to the surface via a stable 1,2,3-triazole linkage. The positions of the

azido and the alkynyl functional groups are interchangeable. The resulting 1,2,3-triazoles are stable at aqueous conditions and high temperature.

In the preferred embodiment of the instant methods, the moiety blocking the 3'-O atom of the dNTP is an allyl moiety and removing it is performed using Na_2PdCl_4 and TPPTS.

This invention also provides a compound comprising a dNTP having bound to its 3' oxygen an allyl or methoxymethyl moiety. In the preferred embodiment, the moiety is an allyl moiety. In another embodiment, the moiety is a methoxymethyl moiety. In specific embodiments the dNTP is dATP, dCTP, dGTP, or dTTP. In a further embodiment, the instant compound is a 3'-O-allyl dNTP, and specifically 3'-O-allyl dATP, 3'-O-allyl dCTP, 3'-O-allyl dGTP, 3'-O-allyl dUTP or 3'-O-allyl dTTP.

Examples of allyl derivatives include, without limitation, analogs or homologs thereof, or haloallyls such as iodoallyl, chloroallyl and fluoroallyl which perform as blocking moieties. Examples of methoxymethyl derivatives include, without limitation, analogs or homologs thereof which perform as blocking moieties.

This invention also provides a process for producing a 3'-O-allyl dNTP comprising:

(a) sequentially contacting a dimethoxytrityl (DMTr) 3' protected nucleoside triphosphate with (i) 3-bromo propene, NaOH and benzene, and (ii) a suitable solvent; and

(b) sequentially contacting the product of step (a) with (i) $POCl_3/(MeO)_3P(O)$, (ii) tributylammonium pyrophosphate, and (iii) $TEAB/NH_4OH$, so as to produce the 3'-O-allyl dNTP.

In one embodiment of the instant method, the suitable solvent of step (a)(ii) is 3% $THF/CHCl_3$. In another embodiment, the concentration of TEAB in step (b)(iii) is about 0.1M.

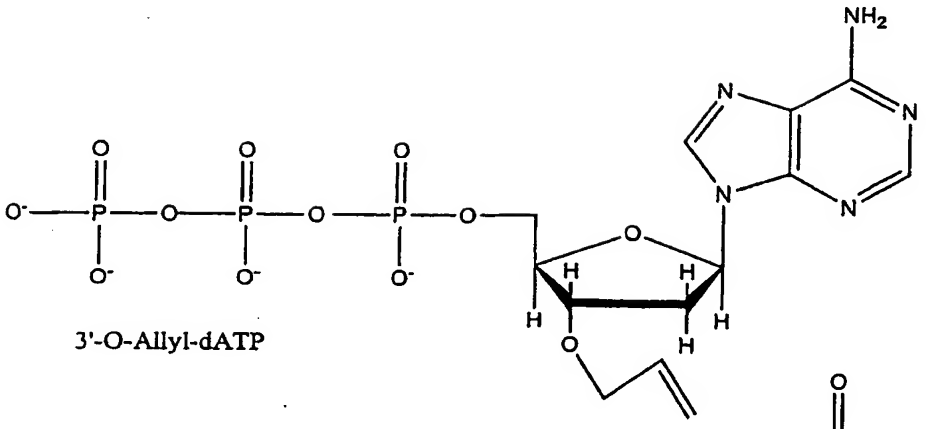
Finally, this invention provides a kit for use in sequencing a single-stranded DNA comprising:

(a) 3'-O-allyl dATP, 3'-O-allyl dCTP, 3'-O-allyl dTTP, and 3'-O-allyl dGTP, each in a separate compartment; and

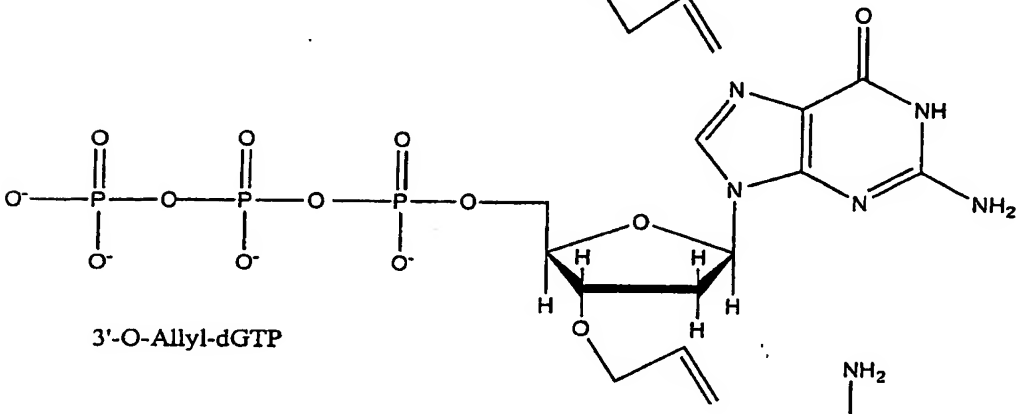
(b) instructions for use.

In various embodiments, the instant kit further comprises (i) a 9°N DNA polymerase (exo-) A485l/Y409V, (ii) reagents permitting DNA polymerization, (iii) reagents permitting pyrophosphate detection using a luciferase-based reaction, (iv) reagents permitting pyrophosphate detection using a coumarin-derived indicator, and/or (v) reagents permitting removal of an allyl group from a 3'-O-allyl dNTP.

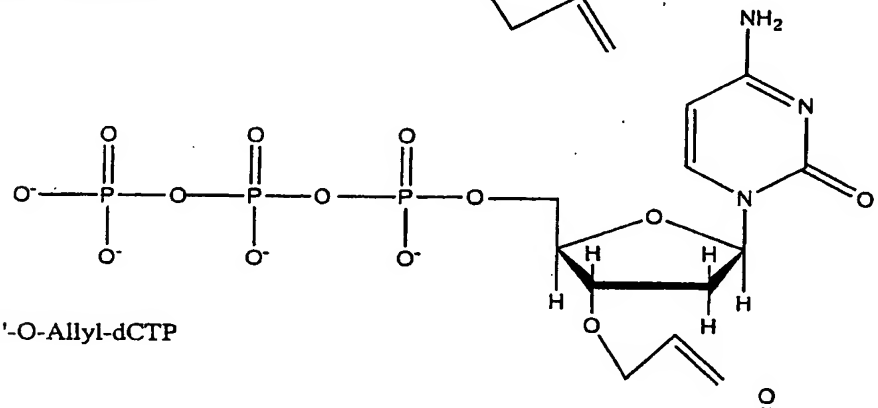
In differing embodiments, the 3'-O-allyl-modified dNTP has one of the following structures:



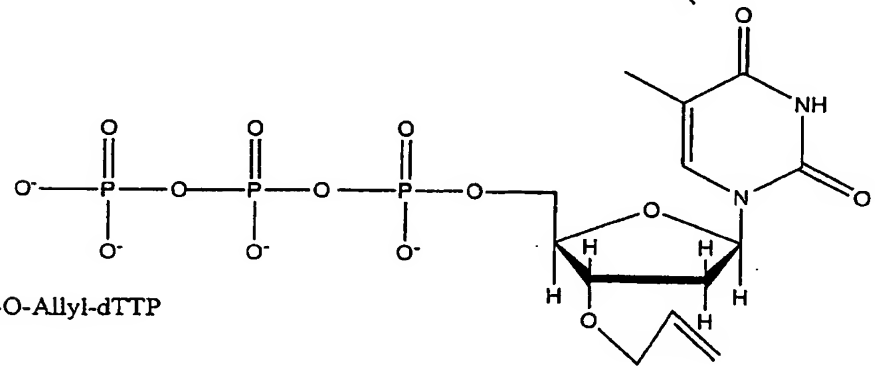
3'-O-Allyl-dATP



3'-O-Allyl-dGTP



3'-O-Allyl-dCTP



3'-O-Allyl-dTTP

This invention will be better understood by reference to the Experimental Details which follow, but those skilled in the art will readily appreciate that the specific experiments detailed are only illustrative of the invention as described more fully in the claims which follow thereafter.

Experimental Details

The general scheme of the improved pyrosequencing method is shown in Fig. 2. 3'-O-allyl-dNTPs (A, C, G, T), instead of dNTPs, are used in the single base extension on a solid surface. Four allyl-dNTPs are added iteratively. Once the complementary base is incorporated, the pyrophosphate that is produced from the reaction can be detected by its reaction with the light-generating luciferase system or a receptor-indicator (R-I) coordination compound via the release of the fluorescent indicator molecule. In one case the R-I compound has negligible or no fluorescence, and the released/displaced indicator is detectably fluorescent. Then the extended primers can be deallylated, washed and reused in the next round. The use of an allyl group solves inherent problems of traditional pyrosequencing.

Synthesis and Deallylation of 3'-O-Allyl-dNTP and its Performance in Single Base Extension

A mild condition to remove a 3'-O-allyl group from DNA in aqueous solution using a catalyst system formed by Na_2PdCl_4 and a water-soluble ligand 3,3',3''-phosphinidynetris(benzenesulfonic acid), trisodium salt (TPPTS) [3] has been identified. Using this condition, the deallylation of the purified 19-mer oligonucleotide (5'-AGAGGATCCAACCGAGAC-T(allyl)-3') (SEQ ID NO:8) was established using MALDI-TOF mass spectrometry. In Fig. 3A, the mass peak at *m/z* 5871 corresponds to the mass of the purified oligonucleotide bearing the allyl group. The deallylation reaction on this oligonucleotide was carried out using the Na_2PdCl_4/TPPTS system. Fig. 3B shows near

complete deallylation with a DNA/catalyst/ligand ratio of 1/50/400 in a reaction time of 30 secs, as shown by the mass peak at *m/z* 5831.

The next step was to prove that the deallylated product could be used in a primer extension reaction and that deallylation did not hinder the continuation of the polymerase reaction. A single base extension reaction using the deallylated product as a primer was performed with a synthetic template and a Biotin-ddGTP nucleotide terminator complementary to the base immediately adjacent to the priming site on the template. The extension product was isolated using solid phase capture purification and analyzed using MALDI-TOF MS [4]. The mass spectrum in Fig. 3C shows a clear peak at *m/z* 6535 corresponding to the extension product proving that the deallylated product can be successfully used as a primer in a polymerase reaction.

These experiments established that Na_2PdCl_4 and TPPTS could be used to efficiently carry out deallylation on DNA in an aqueous environment without the need for an allyl scavenger or harsh conditions. A next step was to ensure that an allyl-modified nucleotide could be incorporated in a DNA Polymerase reaction. For this purpose, a nucleotide analogue 3'-allyloxythymidine triphosphate (3'-O-allyl-dTTP) was synthesized (Fig. 4) and its incorporation ability was tested using a mutant form of 9°N DNA Polymerase (exo-) bearing the mutations A485L and Y409V. Results showed that this enzyme could incorporate 3'-O-allyl-dTTP in a polymerase reaction. 3'-O-allyl-dGTP, 3'-O-allyl-dATP and 3'-O-allyl-dCTP can be similarly prepared according to the scheme set forth in Fig. 4.

The 3'-O-allyl-thymidine triphosphate was used in a primer extension reaction to demonstrate its ability to be incorporated into a growing DNA strand by DNA Polymerase. The extension was performed using a 15-µl reaction mixture consisting of 50 pmol of an 18-mer primer (5'-AGA-GGA-TCC-AAC-CGA-GAC-3') (SEQ ID NO:9), 100 pmol of single-stranded 60-mer DNA template (5'-GTG-TAC-ATC-AAC-ATC-ACC-TAC-CAC-CAT-GTC-AGT-CTC-GGT-TGG-ATC-CTC-TAT-TGT-GTC-CGG-3') (SEQ ID NO:10) corresponding to a portion of exon 7 of the *p53* gene (200 pmol of 3'-O-allyl-thymidine triphosphate), 1X Thermopol reaction buffer (New England Biolabs) and 15 U of 9°N DNA polymerase (exo-) A485L/Y409V. The extension reaction consisted of 15 cycles at 94 °C for 20 sec, 48 °C for 30 sec and 60 °C for 60 sec. The product was desalted using Zip Tip and analyzed using MALDI-TOF MS. The mass spectral data are shown in Fig. 5. Fig. 5(A) shows a single mass peak at *m/z* 5526 corresponding to the unextended primer. Fig. 5(B) shows a single peak at *m/z* 5869 corresponding to the primer extended by a single base 3'-O-allyl-thymidine triphosphate. These data confirm that the above 3'-allyl- modified nucleotide analogue can be efficiently incorporated by 9°N DNA polymerase (exo-) A485L/Y409V.

Single Base Extension on Solid Surface with 3'-Allyl-dNTP (Click Chemistry)

In order to separate primers from the mixture after SBE and deallylation, the primers can be immobilized on a solid surface. One common method is to use paramagnetic beads which are coated with streptavidin. Primers which are labeled with biotin can be attached to the beads because of the biotin-streptavidin attraction. A recently developed

DNA immobilization method using click chemistry, [6] hereby incorporated by reference, can be used in the pyrosequencing method disclosed here. With the addition of template, ally-dNTP and polymerase, the extension can take place on the beads.

A New PPi Detection System for use in Pyrosequencing (Chemosensing Ensemble)

Roger reported a fluorescence chemosensing system (Chemosensing Ensemble) which is described and shown to selectively detect pyrophosphate under physiological conditions [5]. Here, pyrophosphate is capable of displacing a fluorescent coumarin-derived indicator (I) from a bis-Zn^{2+}-dipicolylamine (Zn^{2+}DPA) coordination compound (R). See Fig. 6. With an increase of the Receptor (R) amount, the non-fluorescent R-I coordination compound is formed. When the proportion of R is 50%, the fluorescence reaches its lowest, indicating a 1:1 stoichiometry. Once pyrophosphate is added to the solution, it can replace the Indicator (fluorescence molecule) from the R-I coordination compound. Therefore, the fluorescence molecule is released/displaced, and the fluorescence intensity of the solution will increase (Fig. 7A and B).

Schematically:

$$PPi + R\text{-}I\ Complex \longrightarrow PPi\text{-}R + I$$

(no fluorescence) (strong fluorescence)

Using 3'-allyl-dNTPs as reversible terminators overcomes the inherent problem that the pyrosequencing method otherwise has in accurately detecting the bases in homopolymeric regions, because each base via this invention

is extended one by one with high fidelity. Meanwhile, the newly designed PPi detection system is simple to use and is not affected by dATP. The paradigm of pyrosequencing can be useful in the presence of automatic sequencing machines where each step is repeated in cycles.

Improvement upon the Ronaghi Method

Ronaghi proposed a real time pyrosequencing method in solution [1]. In his method, four enzymes are needed. Among them, sulfurylase is used to transfer PPi to ATP; then luciferase is used to generate light that indicates PPi has been generated. In the next step apyrase is used to degrade ATP and excess dNTP in the reaction; then the process goes to the next round. However, apyrase activity is decreased in later cycles, which is due to the accumulation of intermediate products (such as deoxynucleoside diphosphate, or dNDP) and eventually undegraded dNTP. Because of this limitation, this method can determine the sequence of only about 100 bases at most. See Fig. 8.

However, replacing dNTPs in Fig. 8 with the 3'-Allyl-dNTPs disclosed here, and then following the scheme in Fig. 8, permits one to unambiguously sequence the DNA using repeated cycles without the same degradation problems.

The method disclosed here using the R-I complex can greatly improve the real-time pyrosequencing in the Ronaghi method. (Fig. 9). The R-I complex is used to detect PPi. PPi is converted to PPi-R, while the released indicator I can be transferred to the R-I complex by adding R without removing the components from the solution. The excess dNTP in each cycle is degraded by apyrase. Because there will be no ATP

produced in the detection steps, apyrase now primarily degrades dNTP and is more efficient in its action. Accordingly, more bases can be determined.

Another advantage of this method is that only two kinds of enzymes are used here rather than four, and the detection step will not adversely affect the other steps. However, this improved method cannot detect the bases in homopolymeric regions either, and so 3'-O allyl dNTPs are employed to circumvent this problem.

References

1. Ronaghi M., Uhlen M, Nyren P. A sequencing method based on real-time pyrophosphate. *Science* 281(5375), 363-365 (1998).
2. Ronaghi M., Karamohamed S., Pettersson B., Uhlen M., Nyren P. Real-time DNA sequencing using detection of pyrophosphate release. *Anal. Biochem.* 242(1), 84-89 (1996).
3. "*Design and Synthesis of a 3'-O-Allyloxy Photocleavable Fluorescent Nucleotide as a Reversible Terminator for DNA Sequencing By Synthesis*". H. Ruparel, L. Bi, Z. Li, X. Bai, D. H. Kim, N. Turro & J. Ju. *Proceedings of the National Academy of Sciences USA* 2005, 102, 5932-5937.
4. Edwards, J. R., Itagaki, Y. & Ju, J. Solid Phase Capturable Dideoxynucleotides for Multiplex Genotyping Using Mass Spectrometry (2001). *Nucleic Acids Res.* 29, e104 (p1-6).
5. Roge G. etc. An indicator displacement system for fluorescent detection of phosphate oxyanions under physiological conditions. *Tetrahedron Letters* 45(2004) 8721-8724.
6. Ju, J. et al., U.S. Patent 6,664,079.

What is claimed is:

1. A method for determining the nucleotide sequence of a single-stranded DNA comprising performing the following steps for each nucleic acid residue of the DNA whose identity is to be determined:

(a) contacting the DNA under DNA polymerization-permitting conditions with (i) a 3'-O-blocked dNTP selected from the group consisting of 3'-O-blocked dATP, 3'-O-blocked dCTP, 3'-O-blocked dGTP, and 3'-O-blocked dTTP, and (ii) 9°N DNA polymerase (exo-) A485l/Y409V;

(b) (i) determining whether pyrophosphate is generated as a result of step (a), whereby (1) pyrophosphate generation indicates that polymerization has occurred and the identity of the nucleic acid residue in the DNA is that which is complementary to the 3'-O-blocked dNTP used in part (i) of step (a), and (2) the absence of pyrophosphate generation indicates that the identity of such nucleic acid residue is not that which is complementary to such 3'-O-blocked dNTP, and (ii) if pyrophosphate is not generated, repeating step (a) once, twice or three times as necessary, wherein in each repetition a 3'-O-blocked dNTP is used which is different from any 3'-O-blocked dNTP already used, and determining, after each repetition of step (a), whether pyrophosphate is generated, such generation indicating that polymerization has occurred and the identity of the nucleic acid residue in the DNA is that which is complementary to the 3'-O-

blocked dNTP used in part (i) of the repeated step (a); and

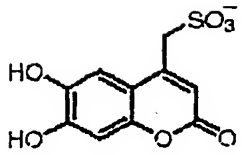
(c) removing from the 3'-O-blocked dNTP polymerized in step (a) or (b), whichever is applicable, the moiety blocking the 3'-O atom of the dNTP, with the proviso that such removing step is optional in the event that there remains no further nucleic acid residue of the DNA whose identity is to be determined.

2. The method of claim 1, wherein determining whether pyrophosphate generated in step (b)(i) is performed by detecting light generated by a luciferase-based reaction.

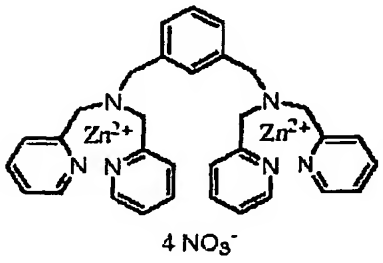
3. The method of claim 2, wherein the luciferase is firefly luciferase.

4. The method of claim 2, wherein the luciferase-based reaction comprises contacting the pyrophosphate with a sulfurylase under conditions permitting the generation of ATP from the pyrophosphate, and contacting the ATP so generated with a luciferase under conditions permitting the generation of light by the luciferase in the presence of ATP.

5. The method of claim 1, wherein determining whether pyrophosphate is generated in step (b)(i) is performed by detecting dissociation of a coumarin-derived indicator from a complex between the indicator and a bis-Zn^{2+}-dipicolylamine coordination compound, wherein the coumarin-derived indicator has the following structure:



and the bis-Zn^{2+}-dipicolylamine coordination compound, when in association with the coumarin-derived indicator, has the following structure:



6. The method of claim 1, wherein the moiety blocking the 3'-O atom of the dNTP is an allyl moiety or a methoxymethyl moiety.

7. The method of claim 6, wherein the moiety is an allyl moiety.

8. The method of claim 1, wherein the DNA is immobilized on a solid substrate.

9. The method of claim 8, wherein the DNA is immobilized on the solid substrate via an azido linkage, an alkynyl linkage, 1,3-dipolar cycloaddition linkage, or biotin-streptavidin interaction.

10. The method of claim 8, wherein the solid substrate is in the form of a chip, a bead, a well, a capillary tube, or a slide.

11. The method of claim 8, wherein the solid substrate is gold, quartz, silica, or plastic.

12. The method of claim 8, wherein the solid substrate is porous.

13. The method of claim 4, wherein the moiety blocking the 3'-O atom of the dNTP is an allyl moiety and the single-stranded DNA is immobilized to a solid substrate.

14. The method of claim 5, wherein the moiety blocking the 3'-O atom of the dNTP is an allyl moiety and the single-stranded DNA is immobilized to a solid substrate.

15. The method of claim 1, wherein the moiety blocking the 3'-O atom of the dNTP is an allyl moiety and removing it is performed using Na_2PdCl_4 and TPPTS.

16. A compound comprising a dNTP having bound to its 3' oxygen an allyl or methoxymethyl moiety.

17. The compound of claim 16, wherein the moiety is an allyl moiety.

18. The compound of claim 16, wherein the moiety is a methoxymethyl moiety.

19. The compound of claim 16, wherein the dNTP is dATP.

20. The compound of claim 16, wherein the dNTP is dCTP.

21. The compound of claim 16, wherein the dNTP is dGTP.

22. The compound of claim 16, wherein the dNTP is dTTP.

23. The compound of claim 16, wherein the compound is a 3'-O-allyl dNTP.

24. The compound of claim 23, wherein the compound is 3'-O-allyl dATP.

25. The compound of claim 23, wherein the compound is 3'-O-allyl dCTP.

26. The compound of claim 23, wherein the compound is 3'-O-allyl dGTP.

27. The compound of claim 23, wherein the compound is 3'-O-allyl dTTP.

28. A compound comprising a dNTP having bound to its 3' oxygen an allyl derivative or a methoxymethyl derivative.

29. A process for producing a 3'-O-allyl dNTP comprising:

(a) sequentially contacting a dimethoxytrityl (DMTr) 3' protected nucleoside triphosphate with (i) 3-bromo propene, NaOH and benzene, and (ii) a suitable solvent; and

(b) sequentially contacting the product of step (a) with (i) $POCl_3/(MeO)_3P(O)$, (ii) tributylammonium pyrophosphate, and (iii) $TEAB/NH_4OH$, so as to produce the 3'-O-allyl dNTP.

30. The method of claim 29, wherein the suitable solvent of step (a)(ii) is 3% THF/$CHCl_3$.

31. The method of claim 29, wherein the concentration of TEAB in step (b)(iii) is about 0.1M.

32. A kit for use in sequencing a single-stranded DNA comprising:
(a) 3'-O-allyl dATP, 3'-O-allyl dCTP, 3'-O-allyl dTTP, and 3'-O-allyl dGTP, each in a separate compartment; and
(b) instructions for use.

33. The kit of claim 32, further comprising a 9°N DNA polymerase (exo-) A485l/Y409V.

34. The kit of claim 33, further comprising reagents permitting polymerization.

35. The kit of claim 33, further comprising reagents permitting pyrophosphate detection using a luciferase-based reaction.

36. The kit of claim 33, further comprising reagents permitting pyrophosphate detection using a coumarin-derived indicator.

37. The kit of claim 33, further comprising reagents permitting removal of an allyl group from a 3'-O-allyl dNTP.

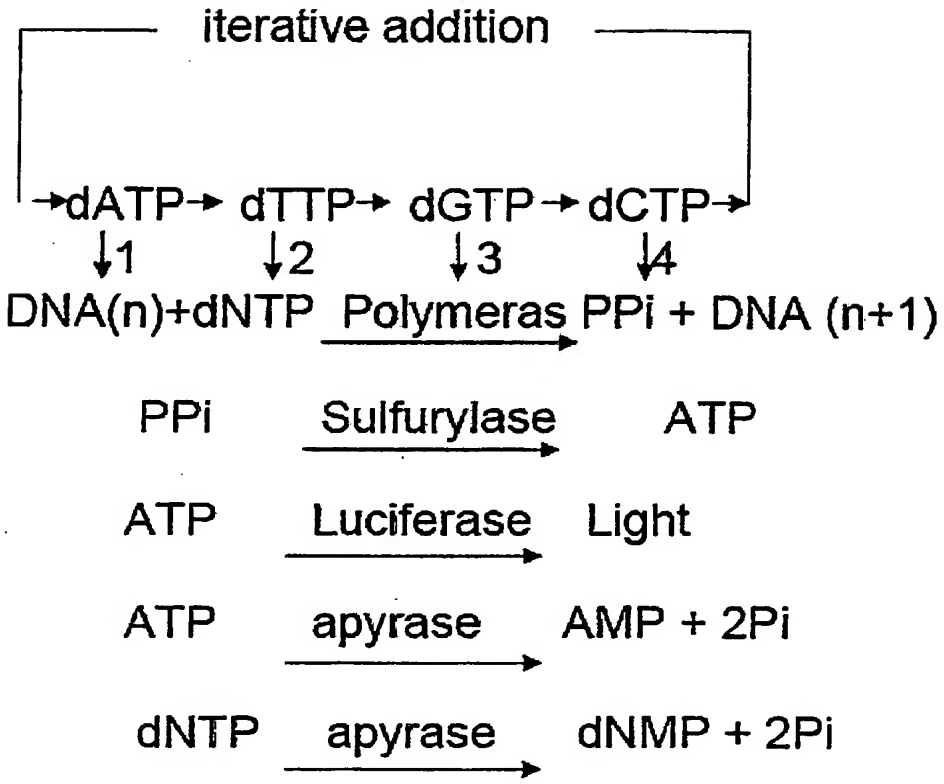
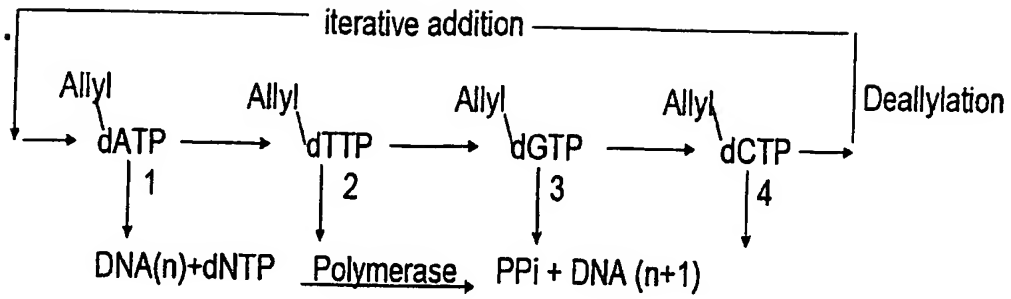

iterative addition
dATP
dTTP
dGTP
dCTP
1
2
3
4
DNA(n)+dNTP Polymeras PPi + DNA (n+1)
PPi Sulfurylase ATP
ATP Luciferase Light
ATP apyrase AMP + 2Pi
dNTP apyrase dNMP + 2Pi

Fig. 1



Non-Fluorescent Fluorescent

Fig. 2

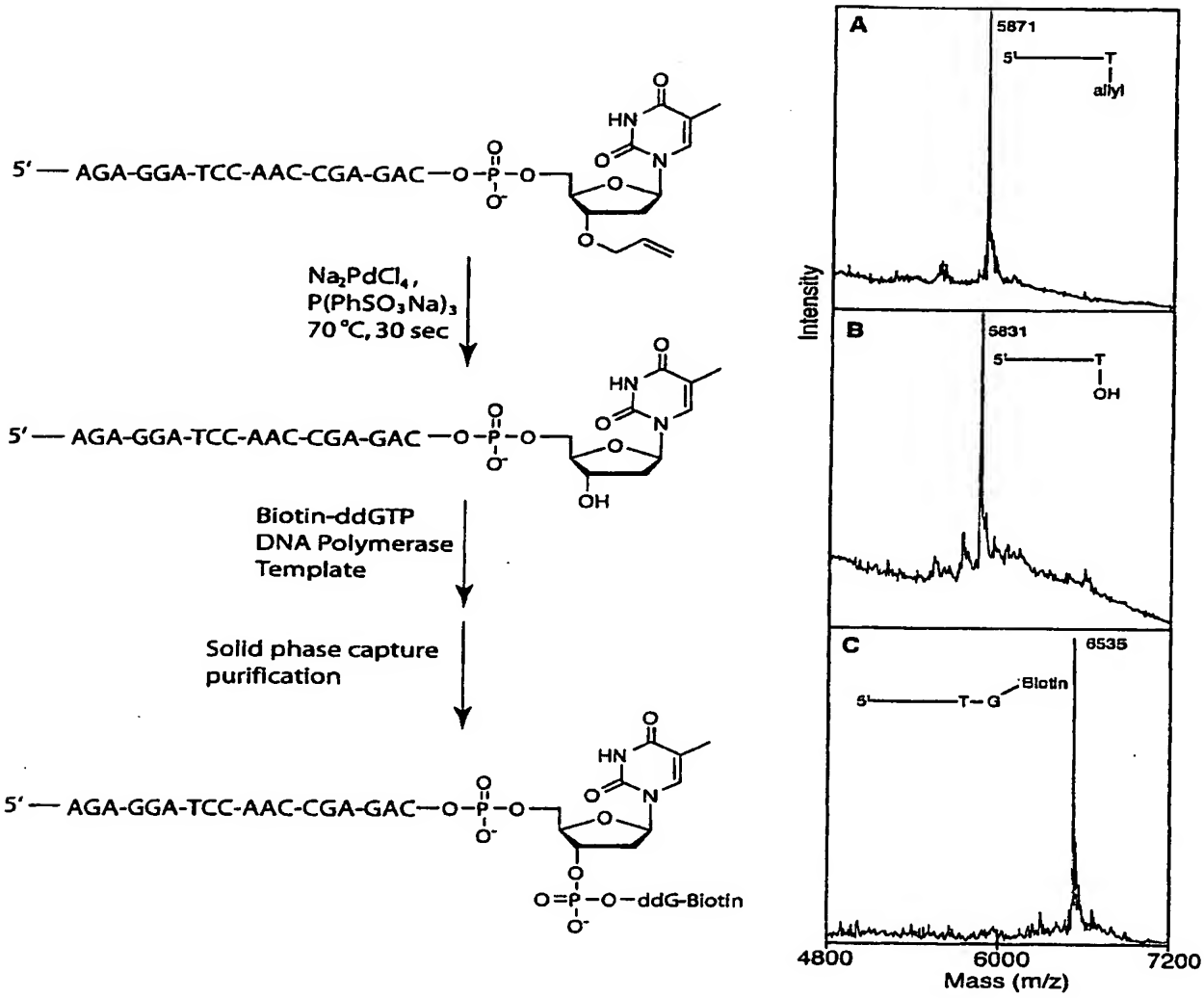

5′ — AGA-GGA-TCC-AAC-CGA-GAC — O
Na2PdCl4,
P(PhSO3Na)3
70 °C, 30 sec
Biotin-ddGTP
DNA Polymerase
Template
Solid phase capture
purification
O=P—O—ddG-Biotin
A
5871
allyl
B
5831
OH
C
6535
Biotin
Intensity
4800
6000
7200
Mass (m/z)

Fig. 3

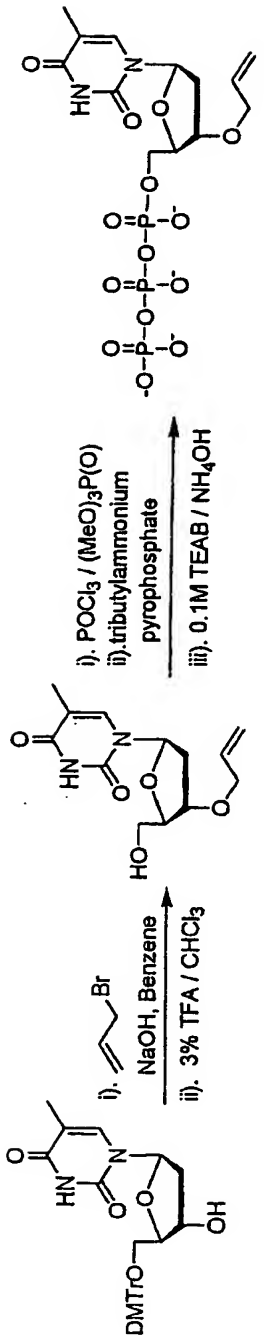

i). $POCl_3$ / $(MeO)_3P(O)$
ii).tributylammonium pyrophosphate
iii). 0.1M TEAB / NH_4OH
i).
Br
NaOH, Benzene
ii). 3% TFA / $CHCl_3$
DMTrO
OH
HO

Fig. 4

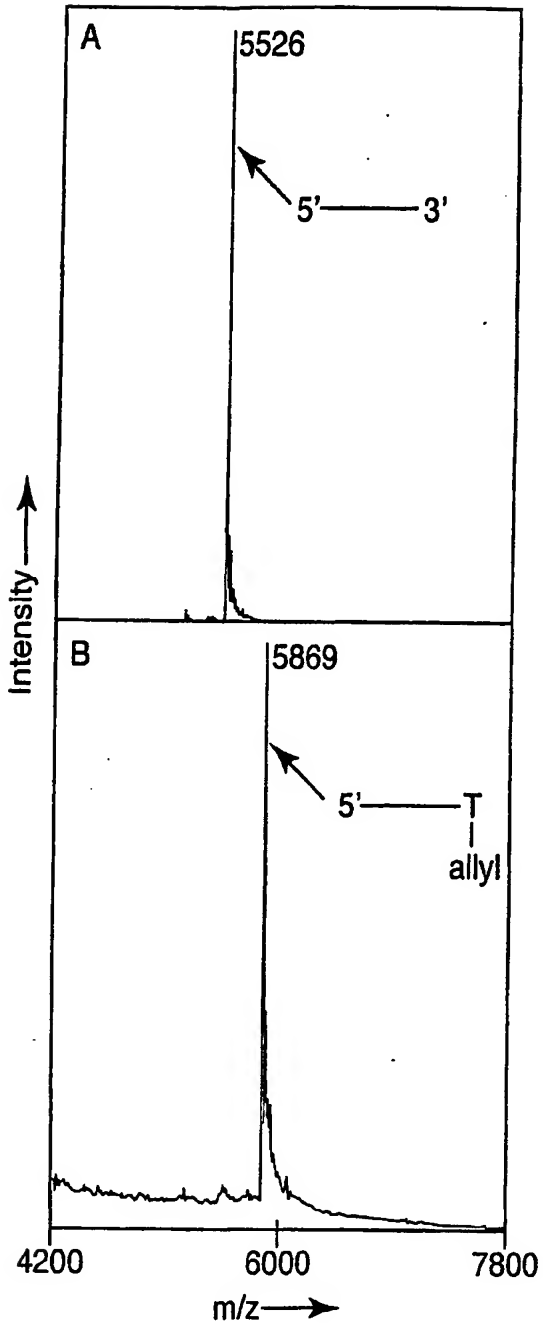
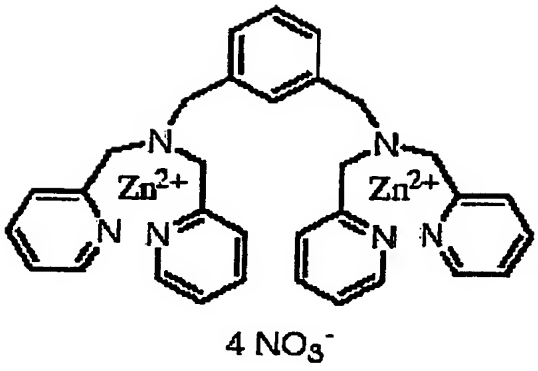


Fig. 5

A.



B.

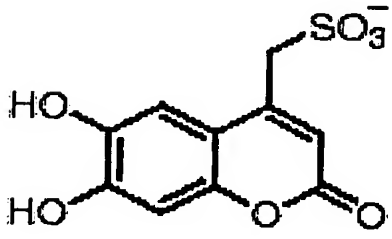
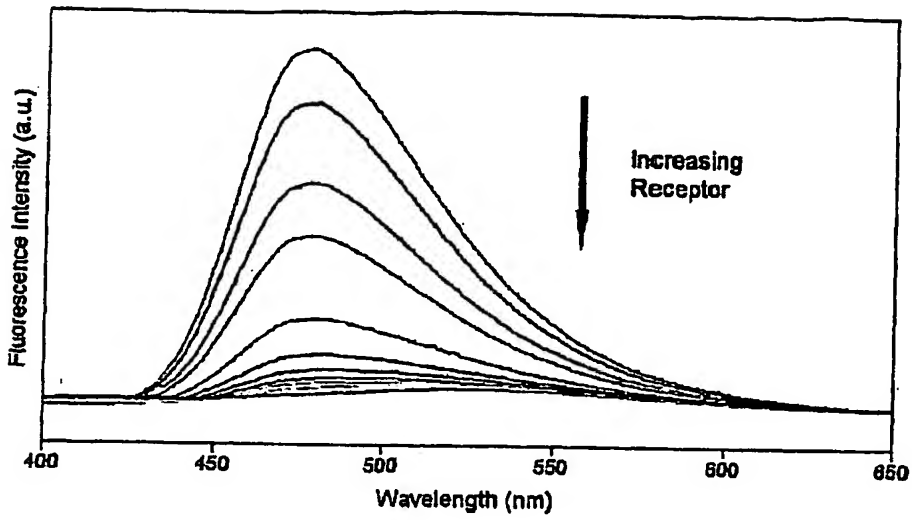


Fig. 6

A.


Fluorescence Intensity (a.u.)
Increasing
Receptor
400
450
500
550
600
650
Wavelength (nm)

B.

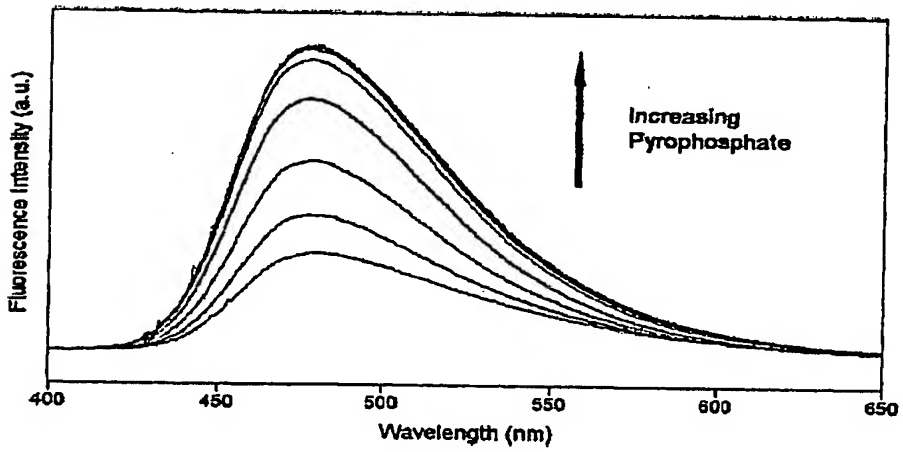

Fluorescence Intensity (a.u.)
Increasing
Pyrophosphate
400
450
500
550
600
650
Wavelength (nm)

Fig. 7

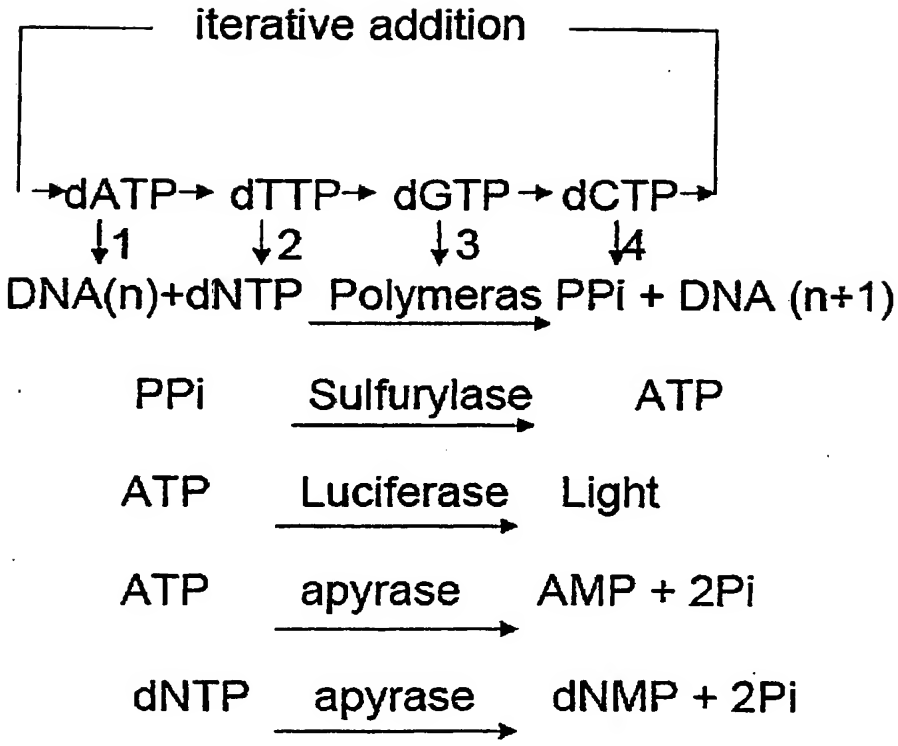

iterative addition
dATP
dTTP
dGTP
dCTP
1
2
3
4
DNA(n)+dNTP Polymeras PPi + DNA (n+1)
PPi Sulfurylase ATP
ATP Luciferase Light
ATP apyrase AMP + 2Pi
dNTP apyrase dNMP + 2Pi

Fig. 8

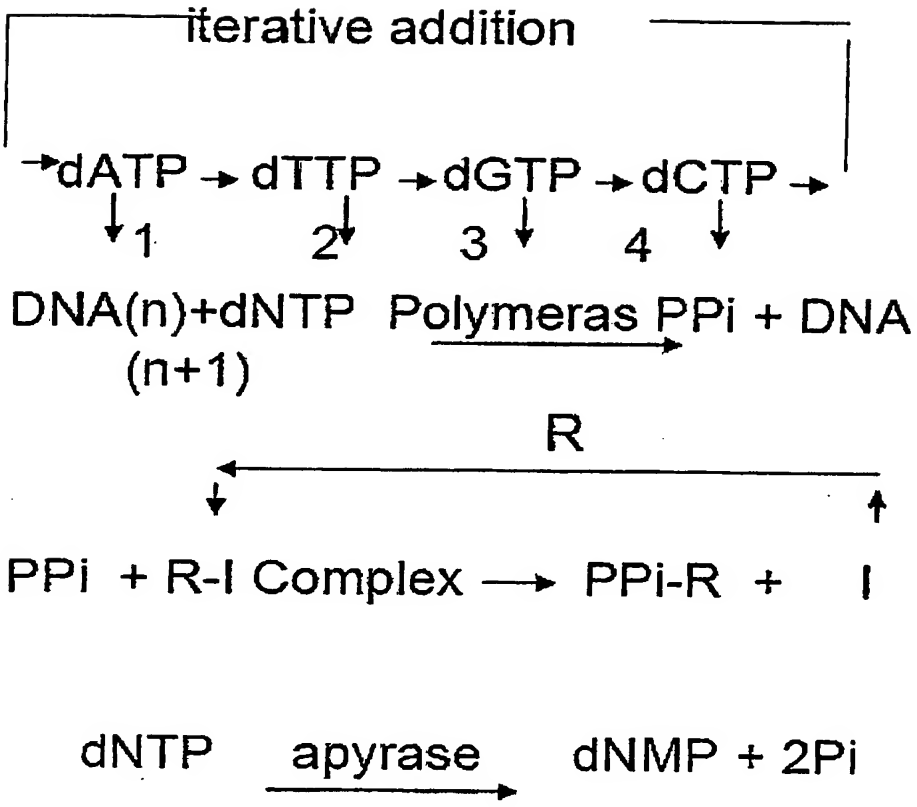
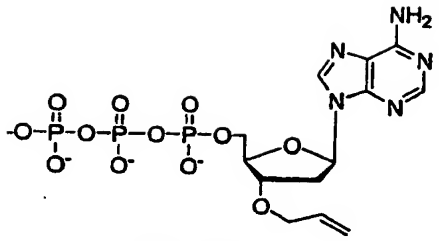
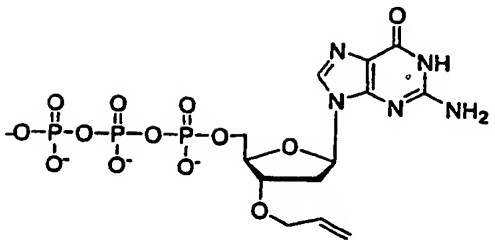

iterative addition
dATP
dTTP
dGTP
dCTP
1
2
3
4
DNA(n)+dNTP
(n+1)
Polymeras
PPi + DNA
R
PPi + R-I Complex
PPi-R +
I
dNTP
apyrase
dNMP + 2Pi

Fig. 9



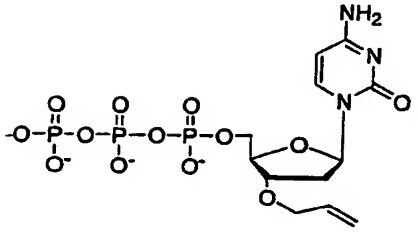
MW=559.07

3'-O-allyl-dATP



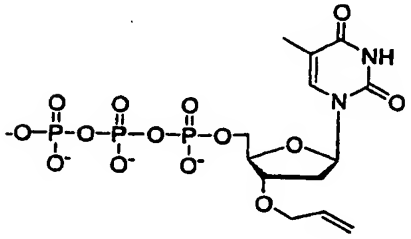
MW=543.00

3'-O-allyl-dGTP



MW=502.99

3'-O-allyl-dCTP



MW=517.99

3'-O-allyl-dTTP

Fig. 10

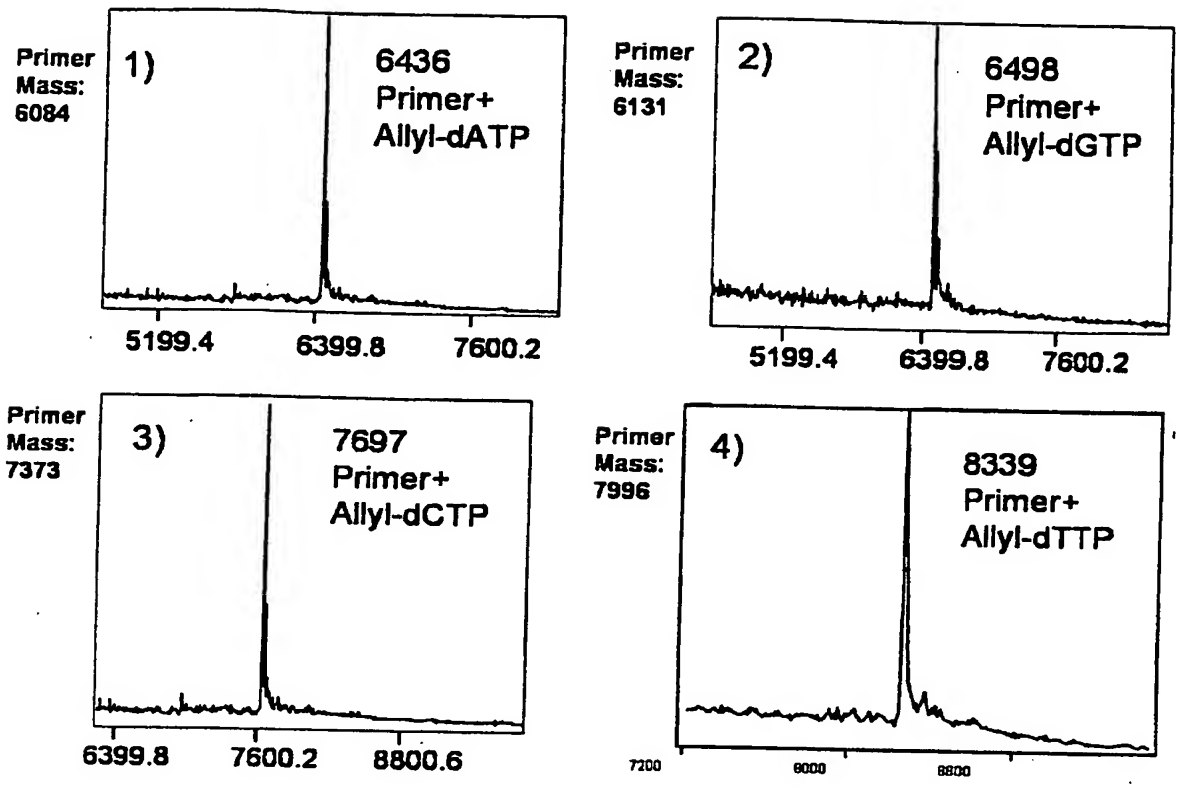

Primer Mass: 6084
1)
6436 Primer+ Allyl-dATP
5199.4
6399.8
7600.2
Primer Mass: 6131
2)
6498 Primer+ Allyl-dGTP
5199.4
6399.8
7600.2
Primer Mass: 7373
3)
7697 Primer+ Allyl-dCTP
6399.8
7600.2
8800.6
Primer Mass: 7996
4)
8339 Primer+ Allyl-dTTP
7200
8000
8800

Fig. 11

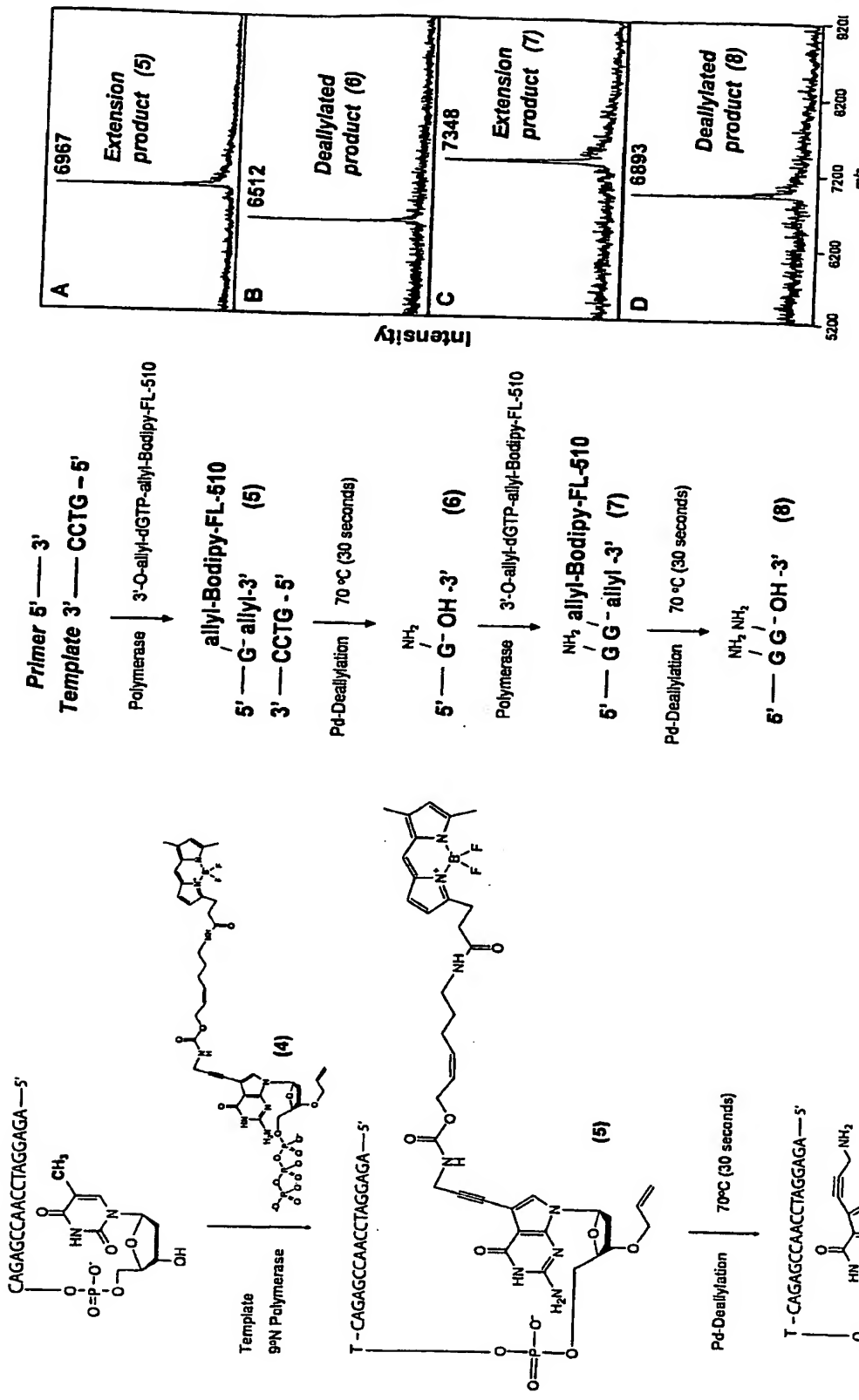

A
6967
Extension product (5)
B
6512
Deallylated product (6)
C
7348
Extension product (7)
D
6893
Deallylated product (8)
Intensity
5200
6200
7200
8200
m/z
Primer 5' — 3'
Template 3' — CCTG – 5'
Polymerase
3'-O-allyl-dGTP-allyl-Bodipy-FL-510
allyl-Bodipy-FL-510
5' — G - allyl-3' (5)
3' — CCTG - 5'
Pd-Deallylation
70 °C (30 seconds)
NH2
5' — G - OH -3' (6)
Polymerase
3'-O-allyl-dGTP-allyl-Bodipy-FL-510
NH2 allyl-Bodipy-FL-510
5' — G G - allyl -3' (7)
Pd-Deallylation
70 °C (30 seconds)
NH2 NH2
5' — G G - OH -3' (8)
CAGAGCCAACCTAGGAGA—5'
Template
9°N Polymerase
(4)
T - CAGAGCCAACCTAGGAGA—5'
(5)
Pd-Deallylation
70°C (30 seconds)
T - CAGAGCCAACCTAGGAGA—5'
(6)

Fig. 12

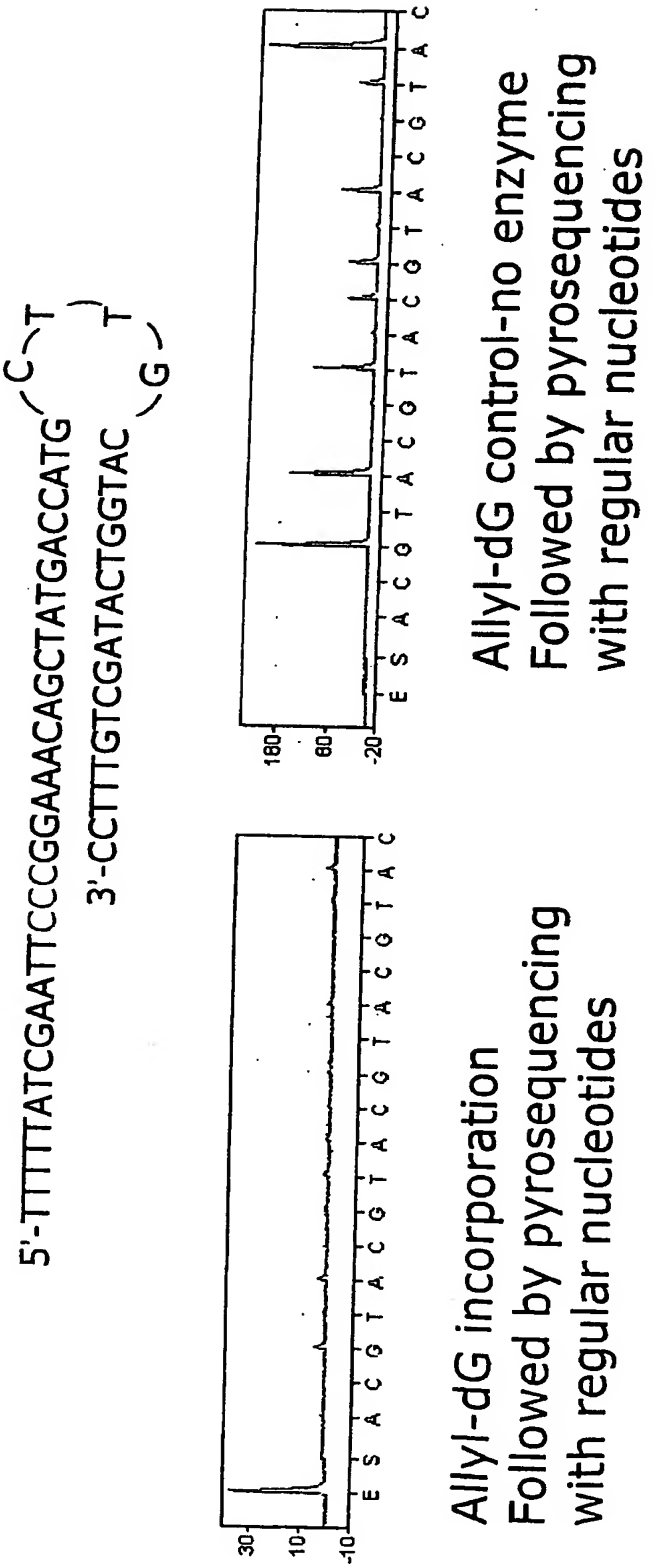

5'-TTTTTATCGAATTCCCGGAAACAGCTATGACCATG
3'-CCTTTGTCGATACTGGTAC
Allyl-dG incorporation
Followed by pyrosequencing
with regular nucleotides
Allyl-dG control-no enzyme
Followed by pyrosequencing
with regular nucleotides

Fig.13

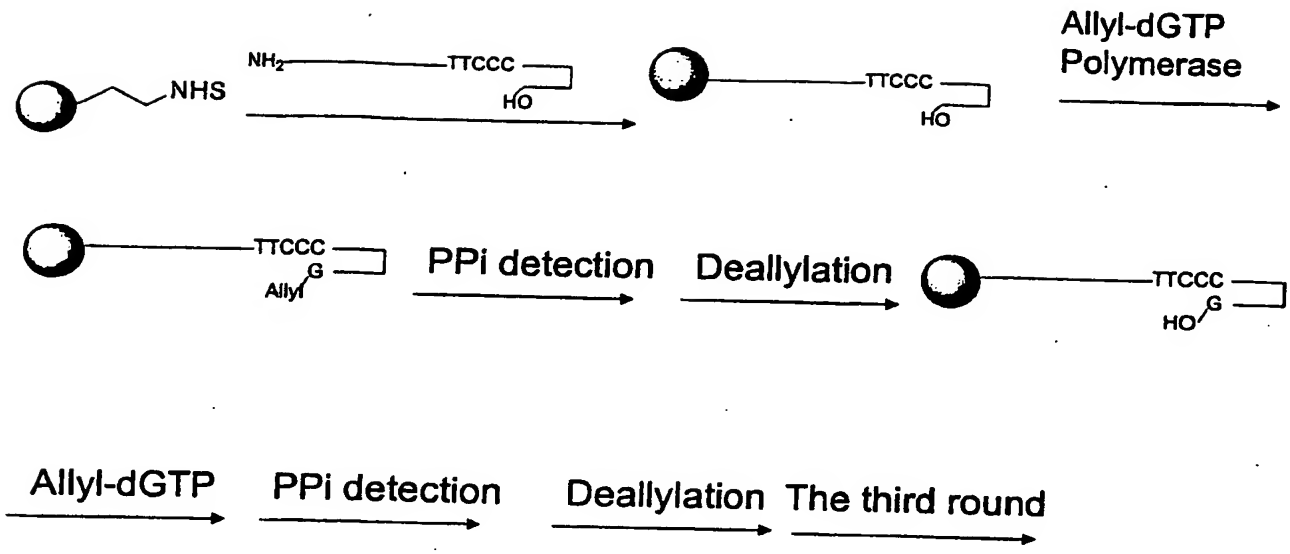
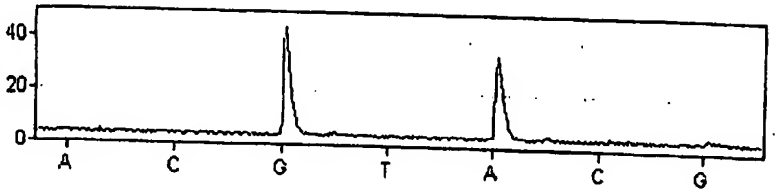
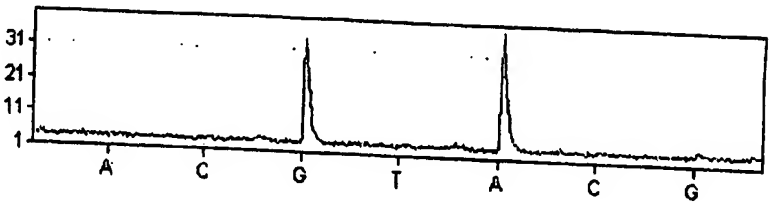

NHS
NH2
TTCCC
HO
TTCCC
HO
Allyl-dGTP
Polymerase
TTCCC
G
Allyl
PPi detection
Deallylation
TTCCC
G
HO
Allyl-dGTP
PPi detection
Deallylation
The third round

Fig. 14

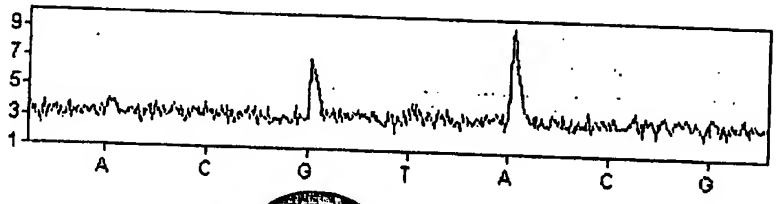


(1) Pyrosequencing results with dNTPs

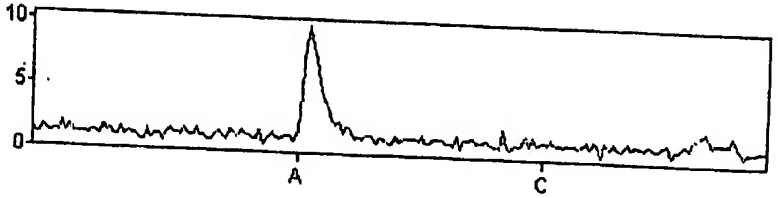


(2) Pyrosequencing results after 1st extension with Allyl-dGTP

(3) Pyrosequencing results after 2nd extension with Allyl-dGTP



(4)Pyrosequencing results after 3rd extension with Allyl-dGTP



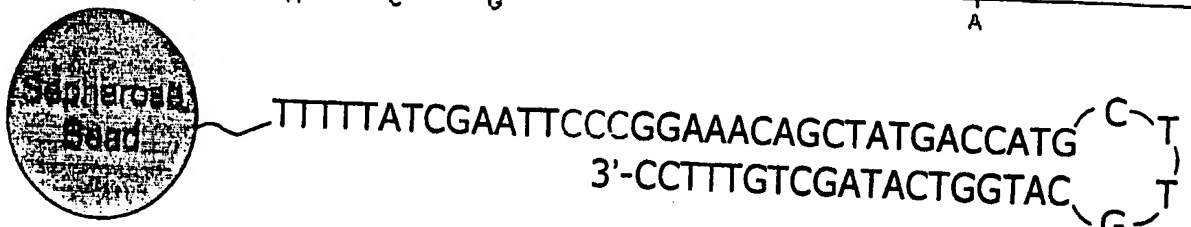
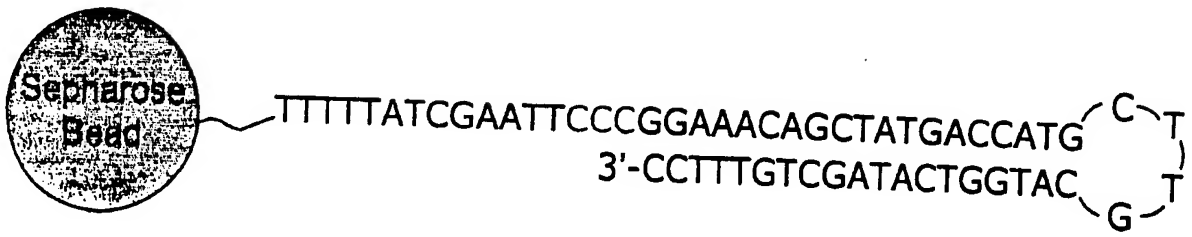
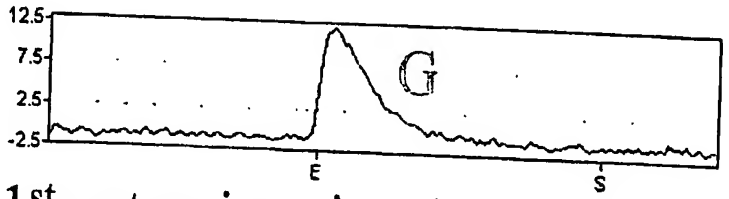
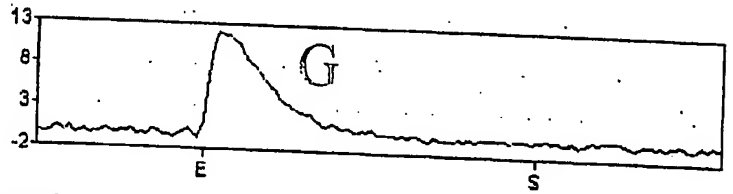


Fig.15

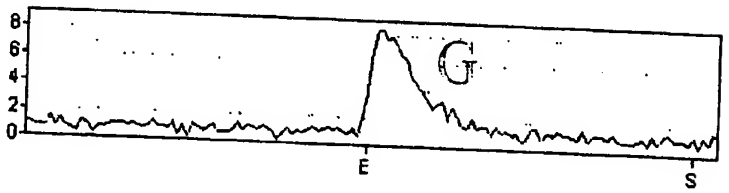




1st extension signal with allyl-dGTP



2nd extension signal with allyl-dGTP



3rd extension signal with allyl-dGTP

Fig. 16

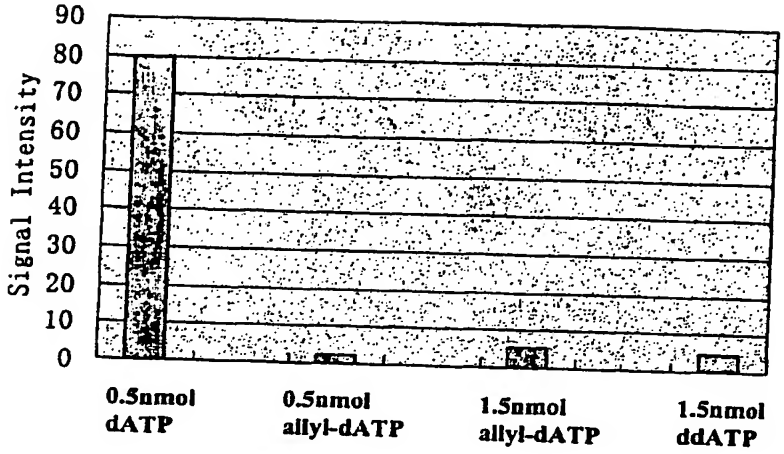
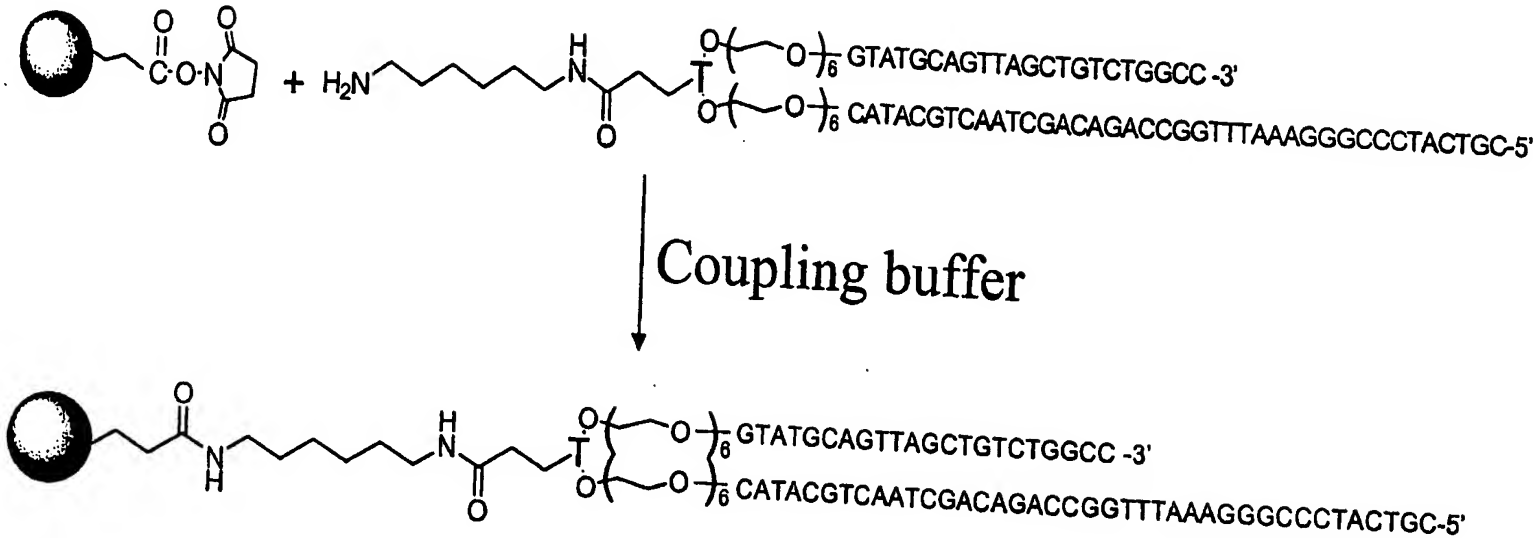

Signal Intensity
90
80
70
60
50
40
30
20
10
0
0.5nmol dATP
0.5nmol allyl-dATP
1.5nmol allyl-dATP
1.5nmol ddATP

Fig. 17



Sequencing the sepharose beads immobilized with DNA with Normal Nucleotides

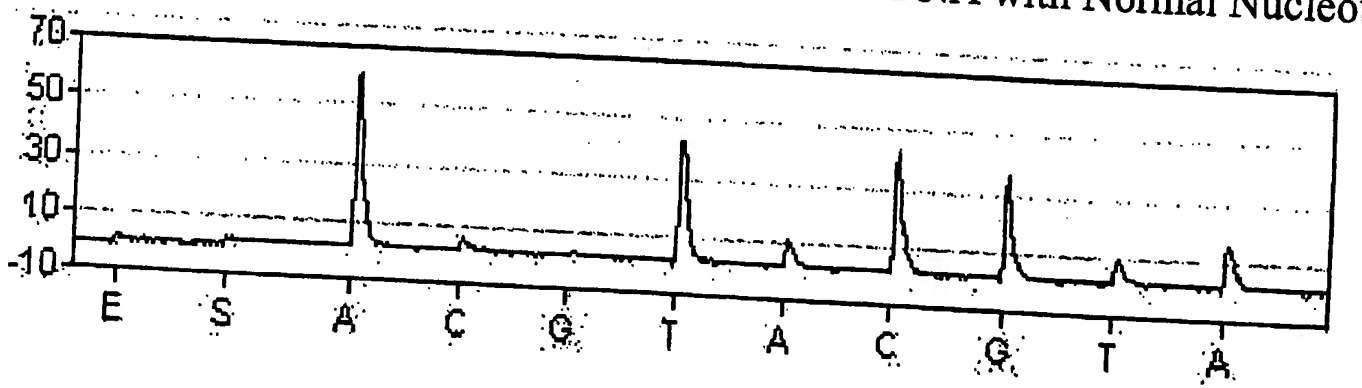
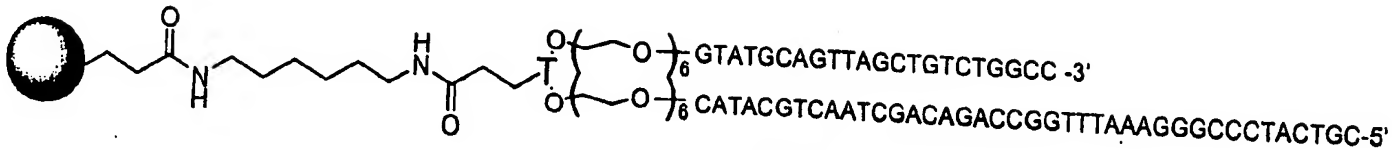
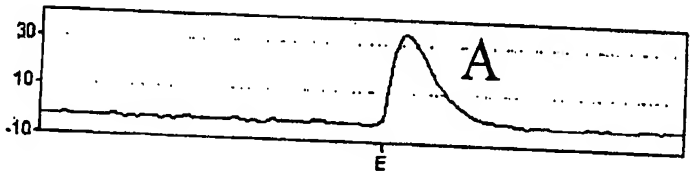


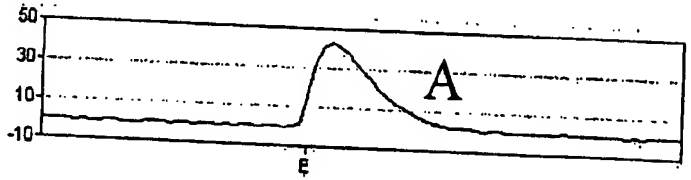
Fig. 18

WO 2007/002204
18/19
PCT/US2006/024157

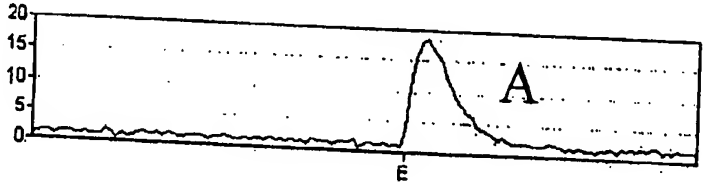




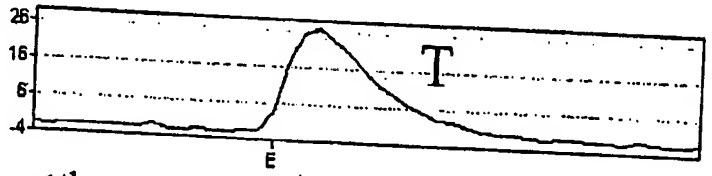
1st -allyl-dATP incorporation



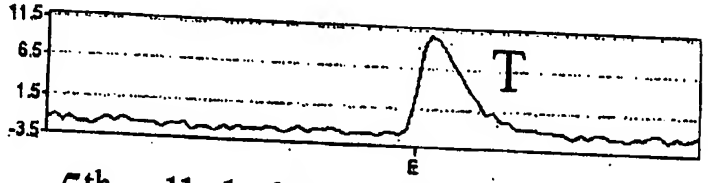
2nd -allyl-dATP incorporation



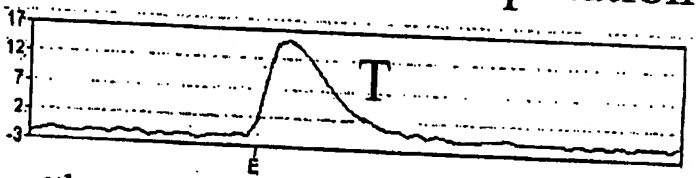
3rd -allyl-dATP incorporation



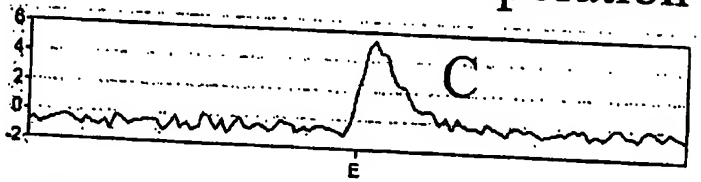
4th - allyl-dTTP incorporation



5th -allyl-dTTP incorporation



6th -allyl-dTTP incorporation



7th -allyl-dCTP incorporation

Fig. 19

WO 2007/002204 PCT/US2006/024157

74689-a-pct.ST25

SEQUENCE LISTING

<110> The Trustees of Columbia University in the City of New York
Ju, Jingyue
Wu, Jian
Kim, Dae H

<120> PYROSEQUENCING METHODS AND RELATED COMPOSITIONS

<130> 0575/74689-A-PCT

<160> 10

<170> PatentIn version 3.3

<210> 1
<211> 18
<212> DNA
<213> Artificial Sequence

<220>
<223> chemically synthesized

<400> 1
agaggatcca accgagac 18

<210> 2
<211> 18
<212> DNA
<213> Artificial Sequence

<220>
<223> chemically synthesized

<400> 2
cagagccaac ctaggaga 18

<210> 3
<211> 58
<212> DNA
<213> Artificial Sequence

<220>
<223> chemically synthesized

<400> 3
tttttatcga attcccggaa acagctatga ccatgcttgc atggtcatag ctgtttcc 58

<210> 4
<211> 5
<212> DNA
<213> Artificial Sequence

<220>
<223> chemically synthesized

<400> 4
ttccc 5

<210> 5
<211> 6
<212> DNA
<213> Artificial Sequence

<220>
<223> chemically synthesized

<400> 5
ttcccg 6

<210> 6
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> chemically synthesized

<400> 6
gtatgcagtt agctgtctgg cc 22

<210> 7
<211> 40
<212> DNA
<213> Artificial Sequence

<220>
<223> chemically synthesized

<400> 7
catacgtcaa tcgacagacc ggtttaaagg gccctactgc 40

<210> 8
<211> 19
<212> DNA
<213> Artifical Sequence

<400> 8
agaggatcca accgagact 19

<210> 9
<211> 18
<212> DNA
<213> Artificial Sequence

<220>
<223> primer directed to human p53 gene

<400> 9
agaggatcca accgagac 18

<210> 10
<211> 60
<212> DNA
<213> homo sapiens

<400> 10

gtgtacatca acatcaccta ccaccatgtc agtctcggtt ggatcctcta ttgtgtccgg 60